DOCUMENT RESUME

ED 380 479 TM 022 785

AUTHOR Kahl, Stuart R.; And Others
TITLE Using Actual Student Work To Determine Cut Scores for Proficiency Levels: New Methods for New Tests.
PUB DATE Jun 94
NOTE 19p.; Paper presented at the National Conference on Large Scale Assessment (Albuquerque, NM, June 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; *Constructed Response; *Cutting Scores; Educational Assessment; Elementary Secondary Education; Language Arts; Mathematics; *Scoring; *State Programs; Student Evaluation; Test Construction; *Testing Programs
IDENTIFIERS Maine; National Assessment of Educational Progress; New Hampshire; *Performance Based Evaluation; *Standard Setting

ABSTRACT

The accountability testing programs of many states have begun to make extensive use of constructed-response questions and to report test results in terms of percentages of students at various proficiency or performance levels. This paper describes the step-by-step procedures for two standard-setting methods recently used for the New Hampshire 1993-94 statewide assessment of language arts and mathematics at grade 3 and the Maine Educational Assessment from the same year, which tested students in grades 4, 8, and 12. Procedures for obtaining cut scores are described. In New Hampshire the Student-based Constructed Response (SBCR) method was used in both language arts and mathematics, and in Maine the SBCR method was used in reading and mathematics and the Item-Based Constructed Response Method was used in all areas. Both of these procedures seem responsive to many of the criticisms leveled at the 1992 achievement levels of the National Assessment of Educational Progress, and both may well be more appropriate than traditional methods. Three tables and seven exhibits provide supplemental information. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

(X) This document has been reproduced as received from the person or organization originating it.

( ) Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
TIMOTHY J. CROCKETT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

USING ACTUAL STUDENT WORK TO DETERMINE CUT SCORES FOR PROFICIENCY LEVELS:

NEW METHODS FOR NEW TESTS

ED 380 479

Stuart R. Kahl
Timothy J. Crockett
Charles A. DePascale
Sally L. Rindfleisch

Advanced Systems in
Measurement and Evaluation, Inc.
Dover, New Hampshire

BEST COPY AVAILABLE

Paper presented at the National Conference on Large Scale Assessment sponsored by the Council of Chief State School Officers, Albuquerque, New Mexico, June 1994.

TM022785

BACKGROUND

The need for new, effective standard-setting procedures for tests relying on constructed-response questions has grown significantly because of two important developments in education. The first is the widespread recognition of the potential negative impacts on curriculum and instruction of testing dominated by the multiple-choice format. The second development is the growing dissatisfaction with norm-referenced reporting of test results because of its failure to convey what it is that students understand and can do. Related to this second issue is educators' increased understanding that seemingly positive normative test results can be inconsistent with students' ability (or lack thereof) to actually perform on more "authentic" and higher order tasks.

As a result of these developments, many states' accountability testing programs have begun to (1) make extensive use of constructed-response (or free-response or open-ended) questions and (2) report test results in terms of percentages of students at various performance or proficiency levels. Such states include (but are not limited to) Delaware, Maryland, Kentucky, Massachusetts, New Hampshire, and Maine. The major purpose of this paper is to describe step-by-step procedures for two standard-setting methods employed recently in New Hampshire and Maine. These methods are refinements of approaches previously used in Massachusetts and Kentucky. The New Hampshire program of interest is the 1993-94 statewide assessment of language arts and mathematics at grade 3. The Maine program is the 1993-94 Maine Educational Assessment (MEA), which tested students in grades 4, 8, and 11 in seven different subject areas. These programs employed both multiple-choice and constructed-response questions, and both used common questions (questions answered by all students in a grade) and matrix-sampled questions (questions unique to different test forms, each student taking only one form). The cut scores identified by the procedures described herein will be used in the reporting of New Hampshire results in the fall and in the reporting of Maine results from the 1994-95 testing.

OVERVIEW OF THE METHODS

In New Hampshire, the Student-Based Constructed Response (SBCR) Method was used in both language arts and mathematics. In Maine, the SBCR Method was used in reading and mathematics, and the Item-Based Constructed Response (IBCR) Method was used in reading, mathematics, science, social studies, and the humanities. (In the latter three areas of the MEA, only matrix-sampled questions were used, each student responding to a limited number of multiple-choice questions and only two constructed-response questions.

A widely used standard setting procedure applied to multiple-choice tests is the Angoff Method. It requires judges to estimate percents correct on multiple-choice questions for borderline students -- i.e., students who are borderline relative to two adjacent proficiency levels. Quite frankly, such estimates are not decisions anyone is qualified to make. The myriad of factors influencing percents correct on multiple-choice items make these judgments little more than sheer guesses. Both the SBCR and IBCR methods require judges to examine actual student work in response to constructed-response questions. Matching student work to predetermined definitions of different proficiency levels is a task virtually anyone is qualified to perform. (The definitions explain what students at various levels within a subject area are able to do.)

The Student-Based Constructed Response (SBCR) Method places students on an IRT (Rasch) ability scale based on their scores on all the "common" questions they answered. Judges review a

complete set of responses for every student whose work they examine. The Item-Based Constructed Response (IBCR) Method places score points for individual items (e.g., the 4-point response to question 1, the 4-point response to question 2, the 3-point response to question 1, etc.) on the IRT (Rasch) ability scale. Judges review responses sorted by score point by item. That is, each folder of responses to be reviewed includes only responses to the same question that earned the same score. Both methods involve the judges in some initial "range-finding" activities, which minimize the number of folders of responses the judges must examine in greater depth. Ultimately, if four levels of proficiency (e.g., distinguished, proficient, apprentice, and novice) are being separated, three cut points on the scaled score continuum must be determined.

There were some minor variations in the SBCR procedures used in New Hampshire and Maine. To avoid confusion, only Maine's procedures for both standard-setting methods are described below. The reader should be aware that responses to all constructed-response questions in the MEA are assigned scores from 0 to 4. Sample performance level definitions and various SBCR and IBCR rating forms are included as exhibits at the end of this paper.

MEA STANDARD-SETTING STEPS

This section describes in detail the steps involved in the SBCR and IBCR methods. Because of the detail, some of the procedures may seem hard to follow on first reading. The reader is urged to refer to the appropriate exhibits attached to the paper as they are discussed in this section.

Meetings:

1. Convene policy advisory committee to create general definitions of proficiency levels. (See Exhibit A.)

2. Convene subject area committees (on-grade teachers, other educators, and non-educators) to translate general definitions of proficiency levels into subject-specific definitions. (See Exhibits B and C.) The abbreviated definitions presented as exhibits will be expanded for release to the field with further explanation and student work samples. These materials are to be consistent with Maine's "Common Core of Learning" and curriculum standards developed by various groups at the national level (e.g., NCTM, AAAS).

3. Convene subject area committees to make judgments for use in standard setting.

Homework:

A complete set of scoring guides must be provided to the judges at Meeting # 2. Before Meeting #3, subject area committees (judges) review open-ended questions and the descriptors of the 4-point (top) responses from the scoring guides. In preparation for the IBCR method, judges should tentatively assign the 4-point responses to either the "distinguished" or the "proficient" category. (The judges use only the scoring guides for this step -- not actual student work.)

Preparations for Student-Based Constructed Response (SBCR) Method for Reading and Mathematics:

1. Produce IRT (Rasch) scaled scores for students based on 5 common questions.

2. Eliminate from the file records of students with highly variable raw item scores, that is, with range greater than 2. (For example, 4,4,2,3,2 is acceptable, but 4,4,3,2,1 is not.)

3. Sample 50 students from each quarter logit. (The students' Rasch ability scores ranged from -2.5 o +2.5 approximately. Thus, there were approximately 20 "quarter-logit" or quarter-unit intervals on that scale.)

4. Rank order students by scaled score.

5. Produce printout listing (in rank order) student name, scaled score, raw scores, lithocode (student serial number), and any other information that would facilitate the location of actual student responses in storage.

6. Identify 10 students in each quarter logit whose response sets are to be pulled: select the 1st student, the last student, and 8 students spaced at equal scaled score intervals in between. (Do not pull responses of students in quarter-logit ranges including students scoring 5 or fewer total raw points (for the 5-item test). Based on the scoring rubrics, students scoring 1 point on the test questions could not be considered above the lowest level of proficiency.)

7. Prepare "homogeneous" folders (one for each quarter logit) each of which includes responses of the 10 students identified in the step above. Place these student response sets in rank order from highest to lowest scaled score in the folder and attach a list of the student lithocodes in the same order to the inside front cover of the folder. Number the outside of the folders consecutively with "1" corresponding to the highest quarter-logit set.

8. Prepare the "heterogeneous" folder which should include copies of the top and the bottom student response sets from every quarter-logit folder. These should be in random order. (Only the leader's heterogeneous folder should list student lithocodes in order by scaled score in the inside front cover.)

9. Produce only a few copies of each homogeneous folder (since judges do not have to examine a particular homogeneous folder at the same time) and one copy of the heterogenous folder for every judge.

10. Prepare SBCR preliminary and final rating forms. (See Exhibits D and E.) The preliminary rating form lists in rank order by scaled score the lithocodes of the students whose response sets are in the heterogeneous folder. The final SBCR rating form is generic.

NOTE: For purely matrix-sampled subject areas in which students answer only two questions, similar procedures for preparing materials would be followed. However, some additional steps could be required. Since each student responded to so few questions, response sets for "virtual" students could be created by merging response sets of students taking different test forms, but matched on ability scores.

Running Meeting 3 - The Standard-Setting Meeting Using the SBCR Method:

1. Provide background, describe procedures, review definitions of proficiency levels. Distribute one heterogeneous folder to every committee member (judge).

2. Ask the judges to locate the work of a subset of students represented in the heterogeneous folder by giving them the lithocodes (in random order) of the top response set in every other homogeneous folder (folder 1, folder 3, folder 5, etc.). (NOTE: These response sets are already in their heterogeneous folders.) Have the judges independently rank order those students' response sets based on overall quality, keeping in mind the proficiency level descriptions. Have the judges record their rank orderings on a small slip of paper. This will not be turned in.

3. Next, write the lithocodes of the response sets just reviewed on newsprint in order from highest to lowest actual performance based on scaled scores. Have the judges note the extent of agreement.

4. Ask the judges to now assign each of the response sets they ranked to a proficiency level. They should each write their decisions on a small slip of paper, again not to be turned in. Record their votes (based on shows of hands) next to the lithocodes on the newsprint.

5. Discuss in depth the response sets just rated as they relate to the proficiency levels definitions. Stimulate discussion with such questions as, "Why did most of you call this student's work 'proficient'?"

6. Have the judges reconsider their ratings of the student response sets and transfer their final ratings to a Preliminary SBCR Rating Form on which the lithocodes of all the response sets in the heterogeneous folder have been entered in order from highest to lowest actual performance.

7. Ask the judges to decide upon the proficiency levels of the rest of the sets in the heterogeneous folder and record their ratings on their preliminary rating forms.

8. Record the "votes" for all response sets on a "master" preliminary rating form based on shows of hands. Then gather the preliminary rating forms.

9. Have the Chief of Standard Setting determine the homogeneous folder or folders that must be evaluated by the judges for determining each of the three cut points. (These would be the folders representing the scaled score intervals in which the transition from one proficiency level to another must occur based on the aggregated ratings from the preliminary rating forms. An example is discussed in a later section.)

10. Divide the group of judges into thirds and have each small group examine the folder or folders for one cut score. Have each judge complete a final SBCR rating form for each folder he/she is assigned. Rotate the materials so that all three small groups examine the folder or folders for every cut point.

Preparations for the Item-Based Constructed Response (IBCR) Method:

1. Determine IRT difficulty/ability associated with each score point from 2 to 4 (inclusive) for all constructed-response items.

2. Prepare the final IBCR rating forms. (See Exhibit G.) The final rating form should be a display placing each score point for each item on the difficulty/ability continuum. A subset of approximately 30 of these score points for items that are fairly evenly distributed over the full ability continuum should be identified by listing them in a separate column on the display.

3. For each of the 30 identified score points for items, prepare a folder containing 20 randomly selected student responses that earned the appropriate score on the item. Identify the score point and the item on the cover of each folder.

4. Prepare the preliminary IBCR rating form (See Exhibit F.) This form lists the same form and item numbers corresponding to the subset of score points identified in steps 2 above, but lists them in the order the items' scoring guides appear in the scoring guide set provided to the judges during Meeting #2 for use in their "homework" assignment.

Running Meeting 3 - The Standard Setting Meeting Using the IBCR Method:

1. Provide background, describe procedures, review definitions of proficiency levels.

2. Ask for shows of hands indicating judges' ratings of the 4-point responses produced as homework, and display frequencies of ratings ("D" or "P") on newsprint.

3. Discuss the items, the ratings, and the descriptions of 4-point responses. Strive for consensus. Also clarify the distinction between score points for items and proficiency levels of students. (e.g., A 4-point response to a question need not correspond to distinguished performance according to the definition of that level. In the end, proficiency levels of students will be based on students' performance on a set of questions collectively. Score points refer only to how an item is scored.)

4. Distribute preliminary IBCR rating forms. Judges should use the scoring guides to complete the preliminary form. They should reconsider homework judgments of 4-point descriptors and also judge the 3-point and 2-point descriptors, recording their judgments on the preliminary rating form. (NOTE: Judges do not need to evaluate all score points for all items -- just those listed on the form.)

5. Collect, by shows of hands, the information from the preliminary rating forms and transfer the aggregated information to a "master" final IBCR rating form displaying the item score points on the ability scale.

6. Discuss cases with widespread ratings (i.e., ratings well distributed over more than two proficiency levels.)

7. Have the Chief of Standard Setting determine the folders that need to be reviewed by judges. For the IBCR method, several folders should be reviewed that represent a probable range in which each cut point will be located.

8. Pass out the final IBCR rating forms. Explain the form to the judges and have them check off the item score points they will be judging.

9. Judges should independently review the responses in the folders and assign to each folder a single proficiency level. It would be best to place the folders for one cut point on a different table from the others. (i.e., Use three tables and have one-third of the judges work on one cut point at a time. However, in the end, each judge must rate the folders for all three cut points.)

10. The judges should force themselves to decide into which of the two proficiency levels in consideration each folder best "fits." These judgments should be recorded on final IBCR rating forms in the spaces to the right of the appropriate score points.

11. Scoring guides for items for which response folders have not been prepared can be used to assist the judges in making their decisions.

12. Once a judge has reviewed the folders in the "vicinity" of each cut point, he/she should estimate a value for each cut point on the numerical scale and record these estimates by drawing arrows at the appropriate places on the numerical scale on the final IBCR rating form. To assist in making their estimates, the judges can look at the scoring guides for the "other" questions.

Using the Judgments to Determine Cut-Points

SBCR — After aggregating the ratings from the SBCR method, it will be clear in which quarter-logit interval or intervals a cut point will be located. Assuming it is one interval for a particular cut point, the aggregated ratings will give us an average proportion of papers in a folder belonging to each of the two proficiency levels under consideration. If four-tenths of the papers are in the upper level, then the cut point would be the scale score within that quarter logit that separates the top four-tenths from the bottom six-tenths of the students within the quarter-logit range. If there is some doubt about which quarter logit "contains" the cut score, then two quarter-logit folders can be merged and the same approach applied to the new half-logit range.

IBCR — The judges work from the IBCR method yields two estimates of cut scores. First, the ratings applied to score points of items will be counted and recorded at the appropriate places on the ability scale display, and then the pattern of entries (such as "16 Ds and 2 Ps") will be examined to determine the most logical points for cut scores. The second estimate for a cut score will be obtained simply by averaging the judges' direct numerical estimates.

NOTE: Cut scores determined by either method can be applied to tests that use multiple-choice items as well, as long as the constructed response and multiple-choice items are scaled together.

EXPLANATION OF SELECTED STEPS OF THE SBCR AND IBCR METHODS

SBCR

Table 1 below shows an aggregation of some of the information from the judges' preliminary SBCR rating forms completed in the standard-setting meetings for reading. These are data from the "range-finding" activity which required the judges to rate student work in the "heterogeneous" folder. The response sets in that folder were the work of the high and low students in each of the ability intervals (.25 units or "logits" on the IRT scale). For each interval, there was a "homogeneous" folder containing the response sets of 10 students (including that interval's "representatives" to the heterogeneous folder). The preliminary ratings depicted in the table led to the identification of folders 2 and 3 as the folders with response sets requiring in-depth examination in order to pinpoint the cut score separating the distinguished (D) and proficient (P) levels.

By picking for the heterogeneous folder the response sets of the high and the low student in each ability interval, we are actually selecting pairs of response sets in which the performance is virtually identical. That is, the low student in one interval performed almost at the same ability level as the high student in the next interval. Thus, we have two indicators at each interval boundary to help determine which homogeneous folders need detailed examination. (NOTE: It is important that the response sets in the heterogeneous folder be ones that were scored very accurately. The computer has only the ratings the scorers assigned to responses to use in placing the students on the ability scale.) Actually, considering the judges' ratings of the low student's work in folder 2 and the high student's work in folder 3, the Chief of Standard Setting could well have decided that only folder 3 needed to be examined.

TABLE 1

Subset of Results from Preliminary SBCR Rating Forms -- Reading

			Frequency of Preliminary Ratings Across Judges			
Folder	Student ID	Location in Interval	D	P	A	N
1	1021048	high	20	-	-	-
1	1121234	low	15	2	-	-
2	1020713	high	11	6	1	-
2	1041031	low	9	7	-	-
3	1051398	high	16	4	-	-
3	1010212	low	1	13	1	-
4	1010596	high	2	15	-	-
4	1120125	low	1	14	1	-
:	:	:	:	:	:	:

Generally, if the scoring of all the work in the different folders is accurate and the students' ability levels fairly accurately determined, then the use of more folders than necessary would have little impact on the final cut point. In this case, the 15 judges' proportions of distinguished students

in folders 2 and 3 combined were: .58, .58, .84, .53, .63, .42, .53, .37, .37, .42, .21, .84, .95, .32, .37. (There were only 19 response sets in the two folders combined instead of the 20 there should have been because a problem response set was rejected from folder 3.) The average of the judges' proportions is .53. Thus, the cut point would be the scaled score that cuts off the top 53 percent of the students in the half-logit interval represented by folders 2 and 3. (That interval happens to be from 2.25 to 2.75 on the ability scale.) The judges' proportions of distinguished students in folder 3 alone were: .56, .33, .89, .44, .33, 0.00, .22, .33, .11, .11, .11, 1.00, .89, .33, .33. The average of these proportions is .40. Thus, the cut point would be the scaled score cutting off the top 40 percent of the students in the quarter-logit interval from 2.25 to 2.50. Since there are relatively few cases in the extreme intervals, the two different cut points would probably be quite close to each other. That is, the point cutting off the top 53 percent of the students in the interval from 2.25 to 2.75 is probably very close to the point cutting off the top 40 percent of the students in the interval from 2.25 to 2.50 because there are relatively few cases in the higher quarter-logit interval.

IBCR

The Table 2 shows aggregated information from the preliminary and final IBCR rating forms completed by judges setting standards in the area of humanities. The subset of the data shown in the table would be used in determining the cut point between the proficient and apprentice levels.

TABLE 2

Subset of Results from Preliminary and Final IBCR Rating Forms -- Humanities

	Frequencies of Preliminary Ratings of Judges				Frequencies of Final Ratings of Judges	
Item/Score Point	D	P	A	N	P	A
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
F5.3-1	1	12	-	-	not reviewed	
F8.3-1	-	11	2	-	not reviewed	
*F1.3-2	-	9	4	-	8	4
*F12.3-1	1	3	9	-	11	1
*F8.3-2	-	6	6	1	7	5
*F5.2-1	-	1	8	4	1	11
*F9.3-1	-	1	12	-	11	1
*F10.3-2	-	4	9	-	3	9
*F11.3-1	-	4	8	1	2	10
F2.2-1	-	-	12	1	not reviewed	
:	:	:	:	:	:	

Recall that the preliminary ratings of score points were based on the judges' review of scoring guides, not student work. This step was completed to minimize the number of folders of

actual student work that had to be examined by the judges. Based on the preliminary frequencies, seven folders were selected for review (identified by asterisks in the table). It appeared that the transition from the proficient to the apprentice levels should occur in the ability scale interval spanned by the score points covered in those folders. (See Exhibit G.) The judges' review of the folders, each of which contained 20 responses corresponding to the appropriate item/score point, confirmed the initial decision. Based on the final ratings (shown in the table above), the cut point would be between item/score points F8.3-2 and F5.2-1 or at approximately 1.4 on the numerical scale.

Notice the problem with the data for item/scale point F9.3-1 in Table 2. Based on the review of the item's scoring guide, the judges almost all believed the 3-point response to question 1 in form 9 matched the definition of the apprentice level for the humanities. However, when the judges reviewed the students' responses, they felt the students' discussions were worthy of a higher rating. Because of the inconsistency of the final ratings for this item/score point with the ratings of other item/score points in the same region of the scale, this score point would be ignored in determining the cut point. This situation is an ideal one in which to refer to the scoring guides for "other" items/score points near F9.3-1 in the ability scale (e.g., F4.3-1 and F12.3-2 in Exhibit G). A similar reversal in judgments occurred for item/score point F12.3-1 in the table.

The last task of the judges was to make their own numerical estimates of cut points. The direct estimates of the proficient/apprentice cut score were: 1.7, 1.0, 1.2, .9, 2.1, 1.5, 1.3, .8, 1.6, 1.5, 1.4, 1.6. The average of these twelve estimates is 1.41 -- almost identical to the cut point one would obtain upon viewing the aggregated ratings of individual items/score points.

DISCUSSION

Recently, the National Academy of Education released a report of a 1993 evaluation of NAEP's 1992 achievement levels entitled, "Setting Performance Standards for Student Achievement." That report (as well as reports of previous studies of NAEP standard setting) was quite critical of (1) the use of the Angoff Method, (2) the inconsistency in judges' ratings, (3) the questionable validity of the cut points, and (4) the questionable validity of achievement level descriptions. The SBCR and IBCR procedures described in this paper seem particularly responsive to many of the specific criticisms in the report of the National Academy of Education. Certainly the methods of standard setting described herein are more appropriate than traditional methods considering the current status of multiple-choice testing.

The judges participating in the MEA standard setting generally felt they were able to relate student responses to definitions of proficiency levels. They felt somewhat less confident in their ability to make judgments based on individual items/score points (the IBCR Method) than they felt using complete sets of responses from students (the SBCR Method). The latter approach is much like the holistic scoring of student portfolios in which many samples of student work illustrate the students' capabilities. Nevertheless, the judges were pleased with the extent of agreement they achieved with respect to various judgments they were asked to make. An added benefit of the procedures is similar to the benefit educators derive from participation in the scoring of student work. In addition to learning some skills that have applications in teaching, the judges found the "true picture" of students' capabilities most enlightening.

Many additional analyses of the data gathered during MEA standard setting will shed additional light on the impact on cut scores of such factors as the method used, the background of judges, and the extent of exposure of the judges to the test questions. The findings of these investigations will be reported during the coming year.

DRAFT DRAFT

MEA PERFORMANCE LEVELS

The MEA performance levels describe the range of performance of students at each grade level assessed. Descriptions of the characteristics of performance levels in each subject will be published as well as these general descriptions.

Distinguished

A distinguished Maine student reveals complete, in-depth understanding of information. The student abstracts the "big ideas" and readily sees long-term as well as short-term implications, parallel situations, and applications and connections of ideas beyond the obvious. This student is able to use insight to communicate complex ideas effectively (and often creatively) and to solve nonroutine problems using innovative, efficient strategies.

Proficient

A proficient Maine student demonstrates the capacity to apply a wealth of knowledge and skills to independently develop new understandings or solutions to routine problems or learning tasks. This student is able to draw important linkages between ideas or procedures and therefore completes tasks and communicates understandings effectively.

Apprentice

An apprentice Maine student displays essential levels of knowledge with partial mastery of higher level concepts and skill application. With occasional coaching, the student can see connections among ideas and sucessfully address problems or learning tasks. This student's communications are direct and reasonably effective, but frequently lack the substance or detail necessary to convey in-depth understanding of concepts.

Novice

A novice Maine student displays partial mastery of essential knowledge and skills. With frequent assistance, the student appears capable of applying understandings to complete well-defined tasks or routine problems. The student's communications are often ineffective and convey only fundamental levels of understanding.

DRAFT DRAFT

MEA PERFORMANCE LEVELS IN READING

The Reading portion of the Maine Educational Assessment (MEA) assesses the readers' ability to communicate their understanding of several different kinds of material text, long and short, taken from various curricular areas, and representing a range of reading levels of difficulty.

Distinguished

A distinguished Maine reader demonstrates an ability to see implications and make applications and connections to ideas beyond the obvious. The student shows insight in understanding complex ideas, control of reading strategies needed to construct meaning from various types of written materials, and knowledge of reference skills.

Proficient (Accomplished?)

A proficient Maine reader demonstrates full understanding and an ability to link ideas within the text and among texts. The students' answers to questions are complete, demonstrate control of reading strategies needed to construct meaning from various types of written material, and show knowledge of reference skills.

Apprentice

An apprentice level Maine reader demonstrates more complete understanding of some types of texts than others. The student may make important connections between ideas within some texts or in some responses but may not be consistent across texts. The reader demonstrates some control of reading strategies needed to construct meaning from various types of written material and knows obvious reference skills.

Novice

The novice level Maine reader demonstrated limited understanding of reading material beyond the obvious stated facts. The student may be able to make connections among ideas stated in some texts but not in others. The reader's control of reading strategies appears to be limited to particular types or difficulty levels of texts. The student may also demonstrate limited ability to use reference skills independently.

DRAFT DRAFT

MEA PERFORMANCE LEVELS IN ARTS AND HUMANITIES

Distinguished

A distinguished Maine student demonstrates a synthesis of elements and principles of composition, a thorough knowledge of subject, clarity of organization, ability to employ original inquiry with expressive qualities to provide creative solutions in his/her responses. He/she demonstrates an in-depth understanding of the connections among the social and historical perspectives and a depth of insight that crosses disciplines. He/she employs multiple viewpoints and creatively analyzes meaning and purpose in terms of experiential connections.

Proficient

A proficient Maine student demonstrates an understanding of elements and principles of composition, knowledge of subject, clear organization, the use of expressive qualities, and appropriate vocabulary to connect ideas and procedures. A clear understanding of major connections among social and historical perspectives is communicated, accurately and analytically with adequate justification of meaning and purpose.

Apprentice

An apprentice Maine student demonstrates an essential understanding of elements and principles of composition, subject, organization and use of expressive qualities. With occasional coaching he/she can see connections among ideas and solve problems. Communication is clear and direct but often lacks detail and originality.

Novice

The novice Maine student displays limited understanding of elements and principles of compositions, subject, organization and use of expressive qualities. With frequent assistance he/she can apply understanding in completing well defined tasks or routine problems. A lack of details, exposure and experience is evident. A partial understanding of connections among the social and historical perspectives is demonstrated.

EXHIBIT D

STUDENT BASED CONSTRUCTED RESPONSE
PRELIMINARY RATING FORM

Judge: _____ Session: A.M. P.M.

Reading - High & Low

1. High 3.16 ID# 1021048 _____
1. Low 3.16 ID# 1121234 _____

2. High 2.74 ID# 1020713 _____
2. Low 2.60 ID# 1041031 _____

3. High 2.48 ID# 1051398 _____
3. Low 2.26 ID# 1010212 _____

4. High 2.24 ID# 1010596 _____
4. Low 2.00 ID# 1120125 _____

5. High 1.88 ID# 1021383 _____
5. Low 1.75 ID# 1021133 _____

6. High 1.73 ID# 1101514 _____
6. Low 1.50 ID# 1040571 _____

7. High 1.44 ID# 1030022 _____
7. Low 1.25 ID# 1110753 _____

8. High 1.22 ID# 1080301 _____
8. Low 1.00 ID# 1050775 _____

9. High .99 ID# 1070899 _____
9. Low .76 ID# 1120555 _____

10. High .74 ID# 1071300 _____
10. Low .50 ID# 1040601 _____

11. High .49 ID# 1021397 _____
11. Low .28 ID# 1110255 _____

12. High .21 ID# 1081552 _____
12. Low .02 ID# 1111522 _____

13. High -.01 ID# 1010784 _____
13. Low -.25 ID# 1120423 _____

14. High -.28 ID# 1011584 _____
14. Low -.50 ID# 1020198 _____

15. High -.52 ID# 1020085 _____
15. Low -.75 ID# 1010403 _____

16. High -.76 ID# 1100147 _____
16. Low -1.00 ID# 1011249 _____

17. High -1.02 ID# 1060409 _____
17. Low -1.25 ID# 1121231 _____

18. High -1.26 ID# 1121713 _____
18. Low -1.50 ID# 1111464 _____

19. High -1.51 ID# 1010296 _____
19. Low -1.73 ID# 1101420 _____

EXHIBIT E

STUDENT BASED CONSTRUCTED RESPONSE RATING FORM

SUBJECT: _____

FOLDER # _____

JUDGE'S NAME: _____ SESSION: AM PM

LITHOCODE #	RATING
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

EXHIBIT F

PRELIMINARY IBCR RATING FORM
HUMANITIES

Judge: _____ Session: A.M. P.M.

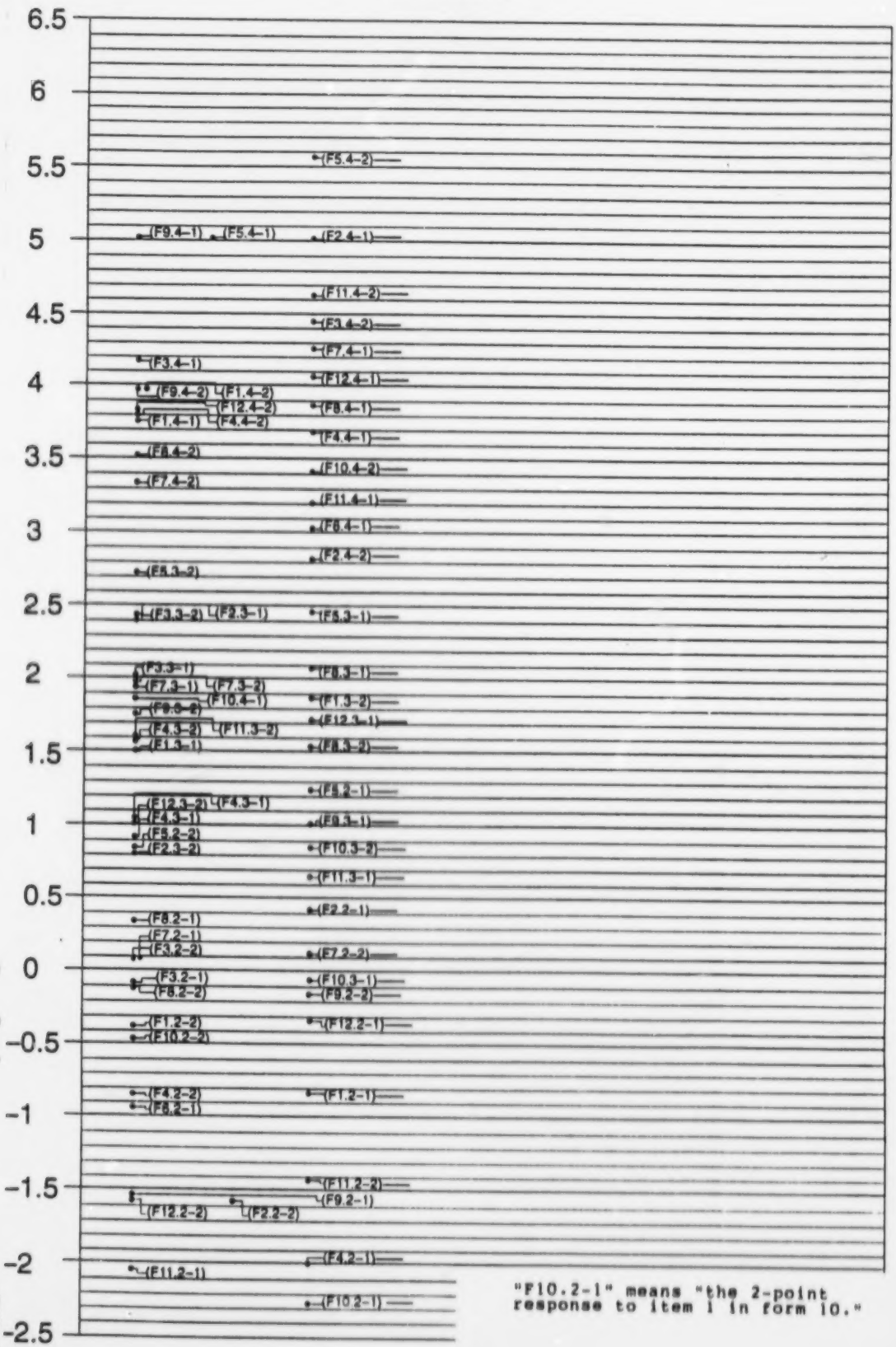
Item	Rating
F1.3.2	_____
F1.2.1	_____
F2.4.1	_____
F2.4.2	_____
F2.2.1	_____
F3.4.2	_____
F4.4.1	_____
F4.2.1	_____
F5.4.2	_____
F5.3.1	_____
F5.2.1	_____
F6.4.1	_____
F7.4.1	_____
F7.2.2	_____
F8.4.1	_____
F8.3.1	_____
F8.3.2	_____
F9.3.1	_____
F9.2.2	_____
F10.4.2	_____
F10.3.2	_____
F10.3.1	_____
F10.2.1	_____
F11.4.2	_____
F11.4.1	_____
F11.3.1	_____
F11.2.2	_____
F12.4.1	_____
F12.3.1	_____
F12.2.1	_____

IBCR RATING FORM

EXHIBIT G

Judge's Name _____

Humanities



"F10.2-1" means "the 2-point response to item 1 in form 10."